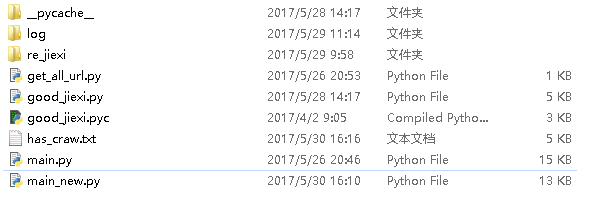
1. 获取店铺所有商品 文件列表



**任务简介**：读取all\_url.csv文件，获取店铺所有商品（product\_list），生成has\_crawl.txt文件

**断点续传：**程序会对比has\_crawl.txt 和all\_url.csv 筛选出为爬的店铺，product\_list设置product\_id 为唯一索引确保商品不会重复

**新的一轮：**需要清空product\_list 表和has\_crawl.txt 文件



配置线程数以及第三方获取IP地址的url

1. 获取商品信息 文件列表

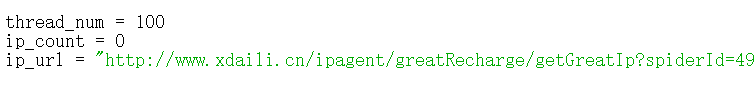


**任务简介**：读取produc\_list 表中的商品ID，获取商品页面的各项信息，生成has\_crawl.txt文件

**断点续传：**程序会对比produc\_list 商品ID 和has\_crawl.txt 中商品ID，筛选出未爬取的商品

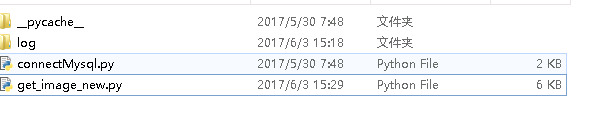
**新的一轮：**需要清空has\_crawl.txt文件

全局变量：



配置线程数以及Ip获取地址

1. 获取图片 文件列表

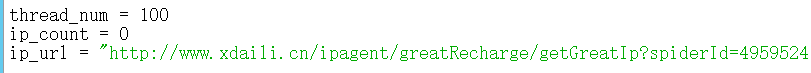


**任务简介：**读取image表 下载图片，将pic\_md5 插入img\_unique 表，

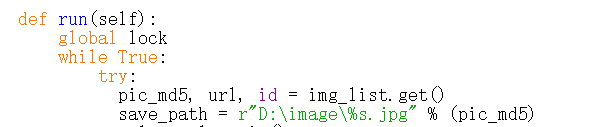
更新image 表isSaved\_Picture 为 1

**断点续传**：所有已经下载的图片pic\_md5会被插入img\_unique 表，其isSaved\_Picture 为 1

全局变量:



配置线程数以及获取IP地址的URL



save\_path 确定下载图片的物理地址